Journal Club

Editor's Note: These short, critical reviews of recent papers in the *Journal*, written exclusively by graduate students or postdoctoral fellows, are intended to summarize the important findings of the paper and provide additional insight and commentary. For more information on the format and purpose of the Journal Club, please see http://www.jneurosci.org/misc/ifa_features.shtml.

The Role of Dorsal Striatal D2-Like Receptors in Reversal Learning: A Reinforcement Learning Viewpoint

Payam Piray
Neuroscience Graduate Program, University of Southern California, Los Angeles, California 90089

Review of Groman et al.

Impaired ability to learn from positive and negative feedback in changing environments is associated with some psychiatric disorders, such as parkinsonism, and impulse control disorders, such as drug addiction and obesity. For example, Parkinson's disease (PD) patients learn better from negative feedback when off PD medication, and some medications reverse this bias, resulting in better learning from positive feedback (Frank et al., 2004). The dopaminergic circuitry is the main target of these diseases (Montague et al., 2004). Recently, some neurocomputational models of these diseases have been proposed based on accumulating evidence that dopamine neurons encode prediction error, i.e., the difference between the obtained reward or punishment and expected value.

Frank et al. (2004) proposed a model to explain the learning deficits associated with PD. They proposed that different types of dopamine receptors within the striatum, especially those in more dorsal regions, mediate the ability to learn from positive and negative prediction error via modulation of dopamine activity in direct and indirect pathways, respectively. According to this hypothesis, positive prediction error increases phasic dopamine release, which results in learning by acting on D1 receptors, whereas negative prediction error results in a dopamine dip below baseline, which results in learning by acting on D2 receptors. The model is based on the reinforcement learning (RL) framework (Frank et al., 2007; Doll et al., 2011). Like in the original RL theory, the model learns the expected value of each option and selects appropriate actions based on these values. The only difference from RL theory is that the speed of learning from negative and positive prediction error can differ. The parameter specifying the speed of learning in RL is called learning rate, which is the degree to which prediction error signal affects subsequent expected value of different choices. In the model proposed by Frank et al. (2007), learning rate for positive errors is related to the availability of D1 receptors within the striatum, whereas the learning rate for negative errors is related to the availability of D2 receptors.

In a recent report, Groman et al. (2011) tested these ideas in nonhuman primates using a task involving reversal learning. The availability of D2-like dopamine receptors in different regions of striatum was measured using positron emission tomography (PET). The task had three phases. First, in the acquisition phase, the subjects chose between three options, only one of which resulted in reward. The monkey learned the rewarding choice by trial and error. Once the performance reached a criterion, the monkey entered the second, retention phase, in which there was no change in stimulus–reward contingencies and the monkey was expected to retain its performance (four correct choices within five consecutive trials). The third, reversal phase, immediately followed, and the previously learned choice was not rewarding anymore and one of the two other choices was rewarding. The subject, again, learned the reward contingencies by trial and error.

The authors measured the relation between performance, the number of trials required to reach criterion in the three phases of the task, and the availability of D2 receptors. While they found no correlation in either acquisition or retention phases, they reported significant correlation between D2 receptor availability in both caudate and putamen and performance during the reversal phase.

The authors next studied the correlation between feedback sensitivity and availability of D2 receptors during the reversal phase. They measured sensitivity to positive feedback by estimating the probability of making a correct response after receiving a positive feedback [Groman et al. (2011), their Fig. 2*A*] and also the probability of perseverative response (choosing the stimulus that was associated with reward before reversal) following positive feedback [Groman et al. (2011), their Fig. 2*B*]. They also measured the sensitivity to negative feedback by estimating the probability of making an incorrect response after receiving a negative feedback [Groman et al. (2011), their Fig. 2*C*]. They found a signifi-

Received June 14, 2011; revised July 13, 2011; accepted July 14, 2011.

I am supported by a Graduate Fellowship from University of Southern California. I am grateful to Michael J. Frank and Nichole Lighthall for reviewing the manuscript.

Correspondence should be addressed to Payam Piray, Hedco Neuroscience Building, 3641 Watt Way, Los Angeles, California 90089-2520. E-mail: ppiray@gmail.com.

DOI:10.1523/JNEUROSCI.3008-11.2011

cant correlation between D2 availability in both caudate or putamen and the probability of choosing a correct response following positive feedback [Groman et al. (2011), their Fig. 2*A*]. They also found a negative correlation between the probability of perseverative response and D2 availability in caudate nucleus, but not in putamen [Groman et al. (2011), their Fig. 2*B*]. They found no significant correlation between D2 receptor availability and the probability of incorrect response after negative feedback [Groman et al. (2011), their Fig. 2*C*].

The correlation between performance and D2 availability [Groman et al. (2011), their Fig. 1] is consistent with the hypothesized role of D2 receptors in learning. During acquisition and retention phases, there is no need to learn from negative prediction error (all possible prediction errors are either positive or zero). However, during the reversal phase, the stimulus that was previously associated with reward was no longer rewarding and so response to this stimulus produces a negative prediction error. Therefore, having more D2 dopamine receptors in dorsal striatum is expected to result in faster learning of the new reward contingencies.

On the other hand, the authors speculated that the correlation between sensitivity to positive feedback and the availability of D2 receptors is opposite to the hypothesized role of D2 receptors in modulating learning from negative feedback. However, this is not actually the case. In Frank et al.'s model (2007), the availability of D2 receptors is modeled with the learning rate for negative prediction errors. According to the model, having more D2 receptors is associated with higher learning rate for negative prediction error, and, correspondingly, faster learning from negative feedback. Therefore, having more D2 receptors more quickly reduces the value of preservative responses. Since three choices were presented simultaneously, reducing the value of preservative responses decreases the probability that it will be chosen and increases the probability that other options, including the correct response, will be chosen. In fact, the model predicts such behavior regardless of the immediately preceding feedback.

To show this formally, I simulated Frank et al.'s (2007) model in the task that monkeys performed. After the acquisition and retention phase (having the probability of correct response >0.8), the model transferred to the reversal phase. Figure 1, *A* and *B*, correspond to Groman et al. (2011), their Figure 2, *A* and *B*, respectively.

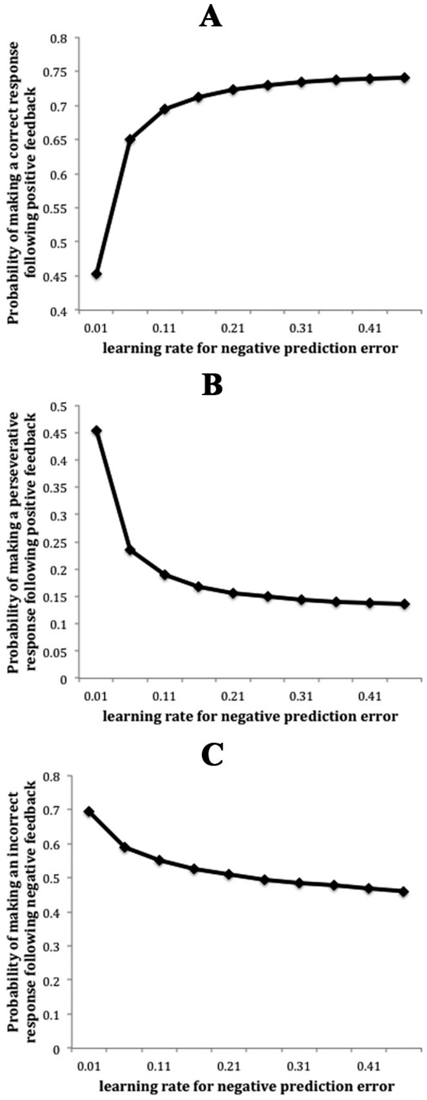

Figure 1. The relation between negative learning rate (associated with availability of D2 receptors) and feedback sensitivity as defined by Groman et al. (2011). The relation between negative learning rate and probability of correct response following positive feedback (***A***), probability of perseverative response following positive feedback (***B***), and probability of an incorrect response following negative feedback (***C***).

The behavior of the model for positive reinforcement is generally consistent with the findings by Groman et al. (2011). Figure 1*A* shows that the probability of making a correct response after receiving positive feedback is higher for those models in which the learning rate is higher with negative feedback (higher availability of D2 receptors). Figure 1*B* shows that the probability of perseverative response following a positive feedback is lower for those models in which the learning rate is higher with negative feedback.

On the other hand, while the model predicts an inverse relation between availability of D2 receptors and the probability of incorrect response following negative feedback (and even after positive feedback) (Fig. 1*C*), the authors found no significant correlation between these two factors [Groman et al. (2011), their Fig. 2*C*]. Although it is surprising in light of Frank et al.'s model, it should be noted that within the model, the effects of higher learning rate for negative prediction error after receiving negative feedback is smaller than its effects after receiving positive feedback, as Figure 1*C* shows. This occurs because trials following negative feedback are more likely to occur very early after reversal, before much learning has occurred. Therefore, overall performance after negative feedback is worse than after positive feedback [which is also true in Groman et al. (2011), their Fig. 2*C*] and so it should be easier to detect an effect after positive feedback than after negative feedback.

In summary, Groman et al. suggest that the lack of correlation between incorrect response and availability of D2 receptors in dorsal striatum during reversal learning in their study is surprising in the light of Frank et al.'s model. But their findings do not support their conclusion that D2 receptors are involved in learning from positive feedback, rather than negative feedback. A more sophisticated method for testing the hypothesized role of D2 receptors would be to use choice data to estimate the learning rates from positive and negative prediction error (Rutledge et al., 2009), and then study the correlation between D2 receptor availability, measured by PET imaging, and learning rates from negative errors.

References

Doll BB, Hutchison KE, Frank MJ (2011) Dopaminergic genes predict individual differences in susceptibility to confirmation bias. J Neurosci 31:6188–6198.

Frank MJ, Seeberger LC, O'reilly RC (2004) By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science 306:1940–1943.

Frank MJ, Moustafa AA, Haughey HM, Curran T, Hutchison KE (2007) Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. Proc Natl Acad Sci U S A 104:16311–16316.

Groman SM, Lee B, London ED, Mandelkern MA, James AS, Feiler K, Rivera R, Dahlbom M, Sossi V, Vandervoort E, Jentsch JD (2011) Dorsal striatal D2-like receptor availability covaries with sensitivity to positive reinforcement during discrimination learning. J Neurosci 31:7291–7299.

Montague PR, Hyman SE, Cohen JD (2004) Computational roles for dopamine in behavioural control. Nature 431:760–767.

Rutledge RB, Lazzaro SC, Lau B, Myers CE, Gluck MA, Glimcher PW (2009) Dopaminergic drugs modulate learning rates and perseveration in Parkinson's patients in a dynamic foraging task. J Neurosci 29:15104–15114.